PURIFICATION OF HISTIDINE TAGGED HIGH-MOBILITY GROUP BOX 3 (HMGB3) EXPRESSED IN *ESCHERICHIA COLI*

Presented by Andrew Chen

In partial fulfillment of the requirements for graduation with the Health Science Scholar's Honors Degree in Biochemistry

Digitally signed by Dalby, Kevin N 2020.05.07 00:07:00 -05'00'

_____ _____

Kevin Dalby, Ph.D
Supervising Professor

Date

May 8, 2020

_____ _____

Jeffrey Barrick, Ph.D
Honors Advisor in Biochemistry

Date

Abstract

High-mobility group box 3 (HMGB3) is a chromatin-associated architectural protein that plays a role in regulating and supporting cellular DNA-dependent processes, such as optimizing DNA replication, transcription, and repair (1). Previous research has shown HMGB3 is often overexpressed in cancerous cells, which makes it a potential, novel target for therapeutic interventions that can help increase the efficacy of existing cancer therapies (2). This study aimed to develop a protocol that can successfully express eukaryotic HMGB3 using the bacterial system *Escherichia coli* (*E. coli*) so that functional HMGB3 can be prepared for high-throughput screenings in order to determine potential inhibitors for future drug studies.

Results revealed that the presence of rare codons within the DNA sequence of HMGB3 were the biggest obstacles to successful expression. After codon optimization, HMGB3 DNA was inserted into a pET-28b vector containing His-tag and expressed alongside GroEL/ES chaperone using the *E. coli* strain BL21 (DE3) competent cells. Purification was achieved utilizing Ni-NTA affinity chromatography and Mono-Q 10/10 anion-exchange chromatography in succession. The purity of HMGB3 protein was then assessed using SDS -PAGE gel.

Introduction

Chromatin is a complex made out of macromolecules that includes DNA and proteins, and it is located within the nucleus of eukaryotic cells. Its primary function is to package long DNA molecules into more compact structures, which helps protect the DNA sequence and regulate DNA dependent processes like the cell cycle. As a result, the structural organization of chromatin is essential for maintaining important gene expressions and cell identity. This structure is highly dynamic and is regulated by chromatin-associated architectural proteins (5). These architectural proteins are part of a central support system for cellular DNA-dependent processes and can regulate the efficiency of DNA repair, transcription, and replication (1). These functional aspects have made chromatin-associated proteins a target of interest for cancer research. Previous studies have shown that cancer-related somatic mutations have been found with high frequency for genes coding for chromatin-associated proteins, and the prognostic outcomes of several cancer types have also been linked to the expression levels of chromatin-associated proteins. Of these proteins, the high mobility group box (HMGB) architectural protein family has been well studied in this regard and is of significant interest (1,5).

HMGB proteins are an abundant and ubiquitous group of DNA binding proteins with a wide variety of cellular functions in the nucleus, cytoplasm, and extracellular matrix; but it is their vital role in regulating DNA damage responses and cell survival following treatment with chemotherapeutic agents, and their association with nucleotide base and strand repair, that makes HMGB proteins a valued target for cancer therapeutics (1, 6). The HMGB family consists of four proteins, with HMGB3 of particular interest due to it often being upregulated in cancer cells and having an important role in the developmental pathway of ovarian cancer (1). Thus, HMGB3 presents a potential, novel target that can significantly enhance current cancer treatments.

In order to identify potential HMGB3 inhibitors via high-throughput screening for drug studies, a bacterial expression system is required to produce functional, recombinant HMGB3 proteins. Current protocols for expressing eukaryotic HMGB3 protein utilize mammalian cells, which can be costly for high-throughput studies and result in unwanted, post-translationally modified proteins (3,4). With no known protocols utilizing bacterial cells to express HMGB3, this study aimed to develop one using *E. coli*.

Materials and Methods

HMGB3 Sequencing and Codon Optimization

Eukaryotic HMGB3 DNA was sequenced by GenScript and its rare codons were optimized for the *E. coli* bacterial expression system using the OptimumGene Codon Optimization software tool. GenScript synthesized a cloning plasmid containing the optimized HMGB3 construct, which would then be transferred to an expression vector.

Construction of a His_6-construct of HMGB3 in pET-28b

The DNA encoding human HMGB3 (GenBank accession number NM_005342) fused to an N-terminal hexahistidine tag was created using the following oligonucleotides: a) forward primer (5'-GGC AGC GGA GGG ATG GCT AAA GGT GAC CCC AAG AAA-3') and b) reverse primer (5'-GAC CCA GAG CCA CC TTA TTC ATC CTC CTC CTC CTC-3'). Following ligation independent protocol, the PCR product and the linearized pET-28b plasmid was digested with T4 DNA Polymerase in order to create complementary sequence overhangs between the insert and vector for ligation. The HMGB3 digestion mix (20μl) contained 16μl of DNA, 2μl of buffer 2.1, and 1μl of dATP and T4 DNA Polymerase. The pET-28b digestion mix (50μl) contained 40μl of DNA, 5μl of buffer 2.1, and 2.5μl of dTTP and T4 DNA Polymerase. The digestion reaction was catalyzed by a thermocycler and the condition was set for 1 hour at 23°C, 20 minutes at 75°C, and storage at 4°C. Then 3μl of the digested HMGB3 PCR product and 1.5μl of the digested pET-28b were mixed and let to anneal at room temperature for 10 minutes before transformation into DH5α *E. coli* cells. The sequence was verified at the Institute for Cell and Molecular Biology (ICMB) Sequencing Facility at the University of Texas at Austin.

Expression and Purification of His-tagged HMGB3

Expression of His-tagged HMGB3

pET-28b-HMGB3 was co-transformed, along with pG-Tf2 (a kind gift from Dr. Gerhard Wagner), which carried the expression system for GroEL/ES chaperone, into the *E.coli* strain BL21 (DE3) competent cells. A single colony of freshly transformed cells was introduced to 10 mL of LB media containing 30 μg/mL of kanamycin and 17 μg/mL chloramphenicol, and grown overnight at 37°C on a shaker (220 rpm). The culture was then diluted 100-fold into TB media containing 30 μg/mL of kanamycin and 17 μg/mL chloramphenicol, and incubated at 37°C (220 rpm). The absorbance level was checked periodically until it reached an OD_{600} of 0.7. Protein

expression was then induced with 50μM isopropyl β-D-1-thiogalactopyranoside (IPTG) for 20 hours at 25°C (220 rpm). The cells were harvested by centrifugation (8000 rpm for 12 min at 4°C) in a 50 mL Falcon tube, flash frozen in liquid nitrogen, and stored at -80°C.

Ni-NTA Affinity Chromatography

The bacterial pellet was thawed on ice and suspended in 300 mL of buffer A (20 mM Tris pH 8.0, 0.03% Brij-30, 0.1% (v/v) β-mercaptoethanol, 5 mM imidazole, 1mM benzamidine, 0.1 mM PMSF, and 0.1 mM TPCK) containing 0.5 M NaCl, 20% glycerol and 0.2 mg/ml lysozyme at 4 °C for 15 minutes, then 15 mL of 10% triton was added and mixed for another 15 min. The resuspended cells were divided into 100 mL fractions on ice and sonicated in a 4°C cold room for 10 min at 5 second intervals with a 50% cycle and 6 microtip limit. The lysate was cleared by centrifugation at 15,000 rpm for 30 minutes at 4°C and the supernatant was divided into 50 mL Falcon tubes. Each tube was introduced with 3.75 mL of Ni-NTA beads resuspended in buffer A, and gently rocked for 1 hour at 4°C. The solution was then transferred to a 100 mL chromatography column where the supernatant was eluted. The beads were then washed with 300 mL buffer A containing 10 mM imidazole and 20% glycerol. His-tagged HMGB3 was then eluted with 30 mL buffer A containing 250 mM imidazole and 20% glycerol.

Mono-Q 10/10 Anion-Exchange Chromatography

The eluted solution was applied to a Mono-Q 10/10 anion exchange column pre-equilibrated with buffer B (20 mM Tris pH 8.0, (*v/v*) 0.03% Brij-30, (*v/v*) 0.1% β-mercaptoethanol, 0.1 mM EDTA, 0.1 mM EGTA, and 10% glycerol). The column was developed over 15-17 column volumes of buffer B with a linear gradient of 0-0.5 M NaCl. The fractions associated with the chromatographic peak were collected and analyzed with SDS-PAGE Gel to confirm the presence and purity of HMGB3. For future experimentation, the fractions were pooled together and dialyzed overnight at 4°C with storage buffer C (25 mM HEPES pH 7.5, 50 mM KCl, 0.1 mM EDTA, 0.1 mM EGTA, 2 mM DTT, and 20 % glycerol).

Results

HMGB3 Sequencing and Codon Optimization

The HMGB3 sequencing and codon optimization results from GenScript are represented by **Fig. 1.** With the OptimumGene Codon Optimization software tool, a total of 139 nucleotides out of 603 nucleotides were altered.

A)

1	ATGGCTAAAG GTGACCCCAA GAAACCAAAG GGCAAGATGT CCGCTTATGC CTTCTTTGTG
61	CAGACATGCA GAGAAGAACA TAAGAAGAAA AACCCAGAGG TCCCTGTCAA TTTTGCGGAA
121	TTTTCCAAGA AGTGCTCTGA GAGGTGGAAG ACGATGTCCG GGAAAGAGAA ATCTAAATTT
181	GATGAAATGG CAAAGGCAGA TAAAGTGCGC TATGATCGGG AAATGAAGGA TTATGGACCA
241	GCTAAGGGAG GCAAGAAGAA GAAGGATCCT AATGCTCCCA AAAGGCCACC GTCTGGATTC
301	TTCCTGTTCT GTTCAGAATT CCGCCCCAAG ATCAAATCCA CAAACCCCGG CATCTCTATT
361	GGAGACGTGG CAAAAAAGCT GGGTGAGATG TGGAATAATT TAAATGACAG TGAAAAGCAG
421	CCTTACATCA CTAAGGCGGC AAAGCTGAAG GAGAAGTATG AGAAGGATGT TGCTGACTAT
481	AAGTCGAAAG GAAAGTTTGA TGGTGCAAAG GGTCCTGCTA AAGTTGCCCG GAAAAAGGTG
541	GAAGAGGAAG ATGAAGAAGA GGAGGAGGAA GAAGAGGAGG AGGAGGAGGA GGAGGATGAA
601	TAA

B)

1	ATGGCGAAGG GTGATCCGAA GAAACCGAAA GGCAAGATGA GCGCGTACGC GTTCTTTGTG
61	CAGACCTGCC GTGAGGAACA CAAGAAAAAG AACCCGGAAG TGCCGGTTAA CTTCGCGGAG
121	TTTAGCAAAA AGTGCAGCGA ACGTTGGAAG ACCATGAGCG GTAAAGAAAA GAGCAAATTC
181	GATGAGATGG CGAAGGCGGA CAAAGTTCGT TACGATCGTG AGATGAAGGA CTATGGTCCG
241	GCGAAAGGTG GCAAAAAGAA AAAGGACCCG AACGCGCCGA AACGTCCGCC GAGCGGCTTC
301	TTTCTGTTCT GCAGCGAATT TCGTCCGAAG ATCAAAAGCA CCAACCCGGG TATCAGCATT
361	GGCGATGTGG CGAAAAAGCT GGGCGAGATG TGGAACAACC TGAACGACAG CGAAAAGCAA
421	CCGTATATTA CCAAAGCGGC GAAGCTGAAA GAAAAGTACG AGAAAGACGT TGCGGATTAT
481	AAAAGCAAGG GCAAATTTGA TGGTGCGAAG GGTCCGGCGA AAGTGGCGCG TAAAAAGGTT
541	GAGGAAGAGG ATGAAGAGGA AGAGGAAGAG GAAGAGGAAG AGGAAGAGGA AGAGGACGAG
601	TAA

C)

MAKGDPKKPKGKMSAYAFFVQTCREEHKKKNPEVPVNFAEFSKKCSERWKTMSGKEKSKFDEM
AKADKVRYDREMKDYGPAKGGKKKKDPNAPKRPPSGFFLFCSEFRPKIKSTNPGISIGDVAKKLG
EMWNNLNDSEKQPYITKAAKLKEKYEKDVADYKSKGKFDGAKGPAKVARKKVEEEDEEEEEEEE
EEEEEEDE

Fig. 1. (A) Original DNA sequence of HMGB3. (B) Optimized DNA sequence of HMGB3 with altered nucleotides highlighted in yellow. (C) Amino acid sequence of optimized HMGB3 (GenBank accession number NM_005342).

Purification of His-tagged HMGB3

The Mono-Q 10/10 anion-exchange chromatography result is represented by **Fig. 2.** The fractions associated with the largest UV peak were collected and run on an SDS-PAGE gel for confirmation as shown by **Fig. 3.** A concentrated band at around 25 kDa in each fraction, along with minimal interfering bands, confirmed the presence of relatively pure His-tagged HMGB3 and indicated a successful expression with the *E. coli* bacterial system.

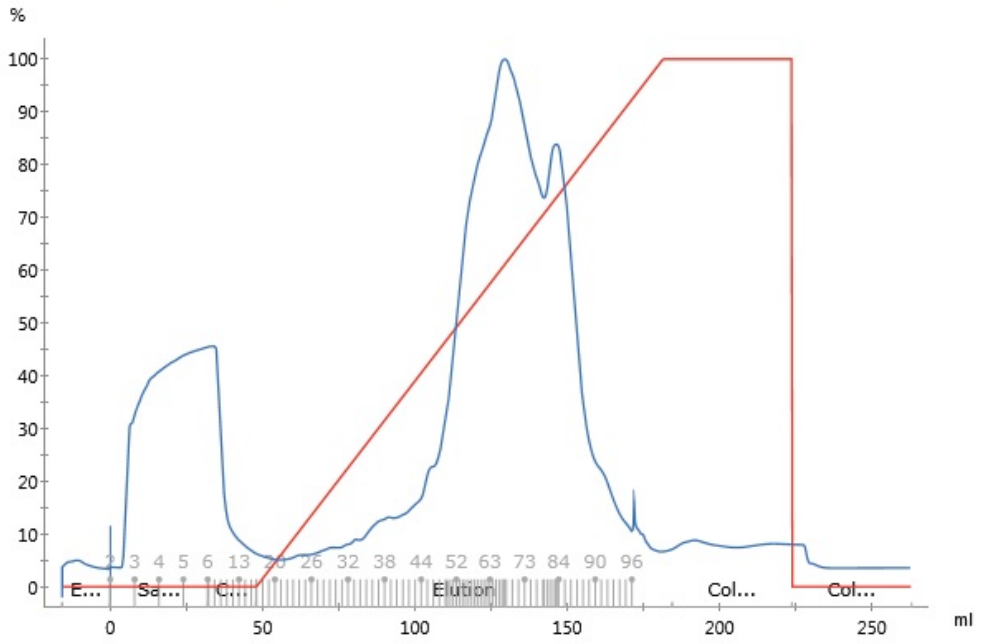

Fig. 2. Eluted His-tagged HMGB3 from Ni-NTA chromatography was purified with Mono-Q 10/10 anion-exchange chromatography. The red line marks the increasing concentration of NaCl while the blue line is ultraviolet signal associated with the presence of proteins.

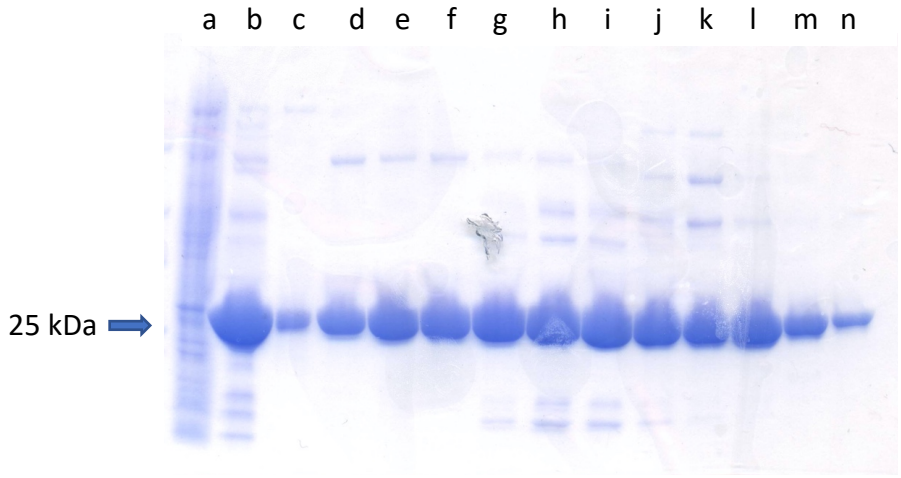

Fig. 3. Samples from the purification process were resolved by SDS-PAGE: bacterial lysate (lane a), Ni-NTA affinity chromatography eluted solution (lane b), Mono-Q 10/10 anion exchange chromatography fractions (lanes c - n).

Discussion

To summarize the findings of this study, His-tagged HMGB3 co-transformed with GroEL-ES chaperone was efficiently expressed in *Escherichia coli* BL21 (DE3) competent cells using a pET-28b expression vector. After the expression was induced for 20 hours with 50 μM IPTG, the protein was isolated from the bacterial cell lysate using Ni-NTA affinity chromatography and further purified with Mono-Q 10/10 anion exchange chromatography (**Fig. 2**). Collecting the fractions associated with the strongest UV signal and running them on SDS-PAGE gel yielded a relatively pure, concentrated band at around 25 kDa, which is associated with His-tagged HMGB3 protein (**Fig. 3**). We found that optimizing HMGB3's rare codons via GenScript's OptimumGene Codon Optimization software tool prior to transformation proved to be the key to successfully expressing eukaryotic HMGB3 with a *E. coli* bacterial system (**Fig.1**).

At the time of this research, there were no known protocol detailing the expression of eukaryotic HMGB3 using a bacterial system. We first investigated several vectors of interest by amplifying their DNA via PCR and checking for sequence stability. The expression vector pET-32a was initially chosen because it contained a thioredoxin tag, which is a fusion tag that helps improve the solubility of the target protein and avoid aggregation of proteins. BL21 (DE3) *E. coli.* cells were initially used due to their high transformation and expression efficiency. However, the SDS-PAGE gel showed an unsuccessful expression. Even with a longer IPTG induced protein expression time, which was altered from 4 hours to 20 hours, the outcome did not change. We hypothesized that BL21 (DE3) *E. coli* cells had difficulty reading the eukaryotic HMGB3 insert due to the presence of rare codons, so Rosetta (DE3) competent cells were used in place of BL21 (DE3) *E. coli* cells because they are designed to enhance the expression of eukaryotic proteins containing codons rarely used in *E. coli*. When the SDS-PAGE gel once again showcased an unsuccessful expression, we analyzed HMGB3 DNA sequences and got its codons

optimized using OptimumGene Codon Optimization software tool to be more compatible with *E. coli*. In addition, we selected to use the pET-28b vector containing only a His-tag, which we hypothesized would better express the optimized HMGB3 construct. Rosetta (DE3) was also replaced with BL21 (DE3) competent cells because the optimized HMGB3 sequence made Rosetta cells redundant, and BL21 cells are better suited to maximize protein expression. Lastly, GroEL/ES chaperone was co-transformed alongside optimized HMGB3 in order to promote optimal protein folding and prevent protein aggregation. Using these conditions, along with a 20-hour IPTG induced protein expression, the SDS-PAGE gel confirmed a successful expression and yielded concentrated bands of relatively pure HMGB3 proteins.

If given more time, the next step would be to assess the activity of the purified HMGB3 protein with enzyme-substrate experiments in order to determine its viability for functional and structural studies. If HMGB3 is proven to be functionally active and structurally sound, then it would be used in high-throughput screenings in order to determine potential inhibitors for future drug studies. This research would have developed a reproducible protocol for preparing viable and pure recombinant HMGB3 via an *E. coli* expression system, and it would be significantly beneficial for studies interested in using HMGB3 as a novel target for cancer therapy.

Acknowledgments

This research was supported by the Dalby Lab, University of Texas at Austin College of Pharmacy. I would like to personally thank Dr. Kevin Dalby and Dr. Tamer Kaoud for their invaluable guidance and mentorship. This project would not have been possible without them.

References

1. Mukherjee, A. and Vasquez, K.M. (2020) Targeting chromosomal architectural HMGB proteins could be the next frontier in cancer therapy. *Cancer Res.* doi: 10.1158/0008-5472.CAN-19-3066
2. Mukherjee, A., Huynh, V., Gaines, K., Reh, W.A., and Vasquez, K.M. (2019) Targeting the High-Mobility Group Box 3 Protein Sensitizes Chemoresistant Ovarian Cancer Cells to Cisplatin. *Cancer Res.* **79**, 3185-3191
3. Wurm, F.M. (2004) Production of recombinant protein therapeutics in cultivated mammalian cells. *Nat. Biotechnol.* **22**, 1393-1398
4. Zhang, Q. and Wang, Y. (2008) High Mobility Group Proteins and Their Post-Translational Modifications. *Biochim. Biophys. Acta.* doi: 10.1016/j.bbapap.2008.04.028
5. Helin, K. and Minucci, S. (2017) The Role of Chromatin-Associated Proteins in Cancer. *Annual Review Cancer Bio.* **1**, 355-377
6. Malarkey, C.S. and Churchill, M.E.A. (2012) The high mobility group box: the ultimate utility player of a cell. *Trends Biochem. Sci.* **37**, 553-562